Language development in mono- and multilingual children: A longitudinal approach

Trebbels, Marina; Duarte, Joana

Language development in mono- and multilingual children: A longitudinal approach

Marina Trebbels, Joana Duarte

Abstract
Both theoretical considerations and empirical results have emphasized the need for longitudinal data in order to gain a more fine-grained insight into the processes of language acquisition and development. Based on data from a two-wave study on the language development of mono- and multilingual children and adolescents in Hamburg (LiMA Panel Study), this article investigates the productive oral language competencies and patterns of language development in 81 children at the beginning of their school careers. The study identifies differences in the patterns of language development in mono- and multilingual children and discusses several challenges that are related to the assessment of oral language competencies in general and in a longitudinal perspective specifically.

Keywords: Language development, Multilingualism, Language competencies, Longitudinal studies

Die Sprachentwicklung ein- und mehrsprachiger Kinder: Eine longitudinale Perspektive

Zusammenfassung
Sowohl theoretische Überlegungen als auch empirische Ergebnisse betonen die Notwendigkeit der Analyse longitudinaler Daten, um tiefere Einblicke in Spracherwerbs- und Sprachentwicklungsprozesse zu gewinnen. Auf Basis einer Studie zur Sprachentwicklung in Hamburg lebender ein- und mehrsprachiger Kinder und Jugendlicher mit zwei Messzeitpunkten (LiMA Panel Study) untersucht der vorliegende Artikel die produktiven oralen Sprachkompetenzen von 81 Kindern zu Beginn ihrer Schulkarriere. Die Studie identifiziert Unterschiede in den Mustern der Sprachentwicklung ein- und mehrsprachiger Kinder und diskutiert mit der Erfassung oraler Sprachkompetenzen, insbesondere in longitudinaler Perspektive, verbundene Herausforderungen.

Schlagworte: Sprachentwicklung, Mehrsprachigkeit, Sprachkompetenzen, Longitudinale Studien

1 Introduction

As a basis for educational support measures, the assessment of language competencies is growingly becoming part of the professionalization requirements of pedagogical staff. Also, language assessment tests have been increasingly employed in large-scale studies that aim to tackle broader questions of educational disparities and inequalities (e.g., PISA, NEPS). Even though this trend has led to an increased development of language assessment instruments (cf. *Reich* 2005), empirical studies most commonly provide results derived from the use of instruments that are focused on concrete language domains or areas (e.g., reading

comprehension, vocabulary, grammar) and involve the measurement of receptive rather than productive competencies (cf. *Jude/Klieme* 2007). Further, both theoretical considerations and empirical results derived from language acquisition studies point to the need for longitudinal data to obtain more fine-grained insight into processes of language development and patterns of language change. Yet, most research within this scope is conducted in a cross-sectional context (cf. *Ortega/Iberri-Shea* 2005; *De Bot/Schrauf* 2009).

While longitudinal studies on language acquisition are scarce in general, an area that is particularly underresearched are processes of language development in school children and adolescents (cf. *Berman* 2004; *Tochinsky* 2004; *Nippold* 2007). Following the consideration that the initial phase of schooling constitutes an important threshold in language development (cf. *Berendes* et al. 2013), this research gap has been recently reversed by a new trend calling for analyses of the "development of competencies relevant to education and participation in social and political life" (*Berendes* et al. 2013, p. 16) of school children. In view of the acknowledgement of a strong link between majority language competencies and educational success on the one hand, and that migrants often grow up in environments that are associated with conditions for language development that strongly differ from those of their native peers on the other hand, a topic that has gained increased attention is the measurement of language competencies in multilinguals (cf. *Eckhardt/Grgic/Leu* 2011).

The present article presents results on the development of the productive oral language competencies in 6 to 7-year-old children based on data that was collected in two measurement points in the context of the LiMA Panel Study (LiPS). After a brief outline of the challenges encountered in the longitudinal measurement of (oral) language competencies, and of the theoretical basis of the empirical study, the article investigates the patterns of development in the productive oral competencies in German of 81 children at the beginning of their school careers and identifies linguistic subdomains that follow different developmental patterns. Following the assumption that language development occurs in comparable stages in mono- and multilinguals (cf. *Clahsen* 1991; *Diehl* et al. 2000; *Reich* 2010), which may yet differ with regard to their length and temporal occurrence (cf. *Mitchell/Myles* 2004), the article further contrasts the patterns of language development in monolingual German speakers with those of their peers with a Turkish, Russian and Vietnamese language background. The article concludes with a discussion of the implications the present results give rise to with regard to the measurement of (oral) language competencies in general and in a longitudinal perspective specifically.

2 Measuring language competencies over time

Each measurement requires a clear theory-based conceptualization of the construct that is to be assessed. The measurement of language competencies is a challenging endeavor not least due to the involvement of several linguistic subdomains that are of very different nature. Educational research makes use of the classic division between productive and receptive competencies on the one hand, and between oral and written language competencies on the other hand (cf. *Jude/Klieme* 2007). Linguistic-based research that specifically addresses issues of language acquisition tends to focus on the measurement of different language components (such as morpho-syntax, semantics and phonology) (cf. *Ehlich* 2007; *Weinert/Ebert* 2013).

Following the complexity of the theoretical construct "language competencies", its measurement over time appears to be even more challenging. Indeed, the literature provides very few models only that aim at conceptualizing language competencies for longitudinal measurement purposes. One approach describes language development as a non-linear process which occurs in stages leading to certain milestones that differ across linguistic subdomains (cf. *Mitchell/Myles* 2004; *Karmiloff-Smith* 1995). Attempting to conceptualize language competencies for purposes of longitudinal measurement, *Ehlich/Bredel/Reich* (2008) identify six broad language domains that are subject to substantial change over time: (1) phonic skills that comprise aspects of pronunciation and intonation; (2) the semantic dimension relating to vocabulary acquisition and concept development; (3) the morpho-syntactic dimension addressing grammar rules and sentence connecting aspects; (4) pragmatic skills that involve interactional and situational aspects of communication; (5) the discursive dimension relating to the ability of linguistic target-specific interaction with others; and (6) the literary aspect involving all competencies related to writing skills, orthography and textuality. According to the authors (cf. *Ehlich/Bredel/Reich* 2008, p. 15), the clear distinction between aspects relevant to younger children and those who are more marked in older speakers allows the identification of broad language areas to be covered when investigating language competencies and change at different ages.

Given the existence of several linguistic subdomains that need to be considered in the assessment of language competencies, a major challenge related to the measurement of language development is to ensure the equatability of observations over time (cf. *Ortega/ Iberri-Shea* 2005). A first central question, which has remained unresolved in the literature, is whether it is more appropriate to use the same instrument multiple times or to use different instruments in order to measure the same linguistic construct over time (cf. *Bachmann* 2007). On the one hand, using the same instrument may increase the comparability of observations over time as variations in the instrument pose the difficulty of disentangling time-induced and topic- or task-induced variation in language competencies. On the other hand, the concern has been voiced that results derived from the repeated use of the same instrument are more vulnerable to test effects that result from aspects such as lower levels of motivation or boredom (cf. *Ortega/Iberri-Shea* 2005). Further, the use of the same instrument is more likely to cause ceiling effects that prevent a closer discrimination of higher abilities (cf. *Bachmann* 2007).

As regards the collection of oral language data in particular, concerns have been voiced that interviewer effects may further compromise the validity of test instruments (cf. *Shohamy* 1983; *Ross/Berwick* 1992; *McNamara* 1996; *O'Loughlin* 1997; *Brown* 2003). The literature discusses a variety of effects that relate to the proficiency of interviewers, ranging from speech accommodation effects (e.g., interviewers adjust their speech to the participant, cf. *Ross/Berwick* 1992; *Malvern/Richards* 2002), to variation in the interviewers' language proficiency and their capacity to master the interviewing procedure (e.g., the extent to which interviewers adhere to the interview script, cf. *McNamara/Lumley* 1997) and their personality and communication style (e.g., with regard to questioning techniques and the type of feedback provided to encourage participants; see for example *Brown* 2003; *Ross* 1996; *Brown/Lumley* 1997).

A further consideration relates to the measurement of language development in multilinguals. While it has been suggested that language development in mono- and multilingual children occurs in comparable stages, the literature provides evidence that these stages may strongly differ both in length and temporal occurrence across mono- and mul-

tilinguals (cf. *Clahsen* 1991; *Karmiloff-Smith* 1995; *Diehl* et al. 2000; *Mitchell/Myles* 2004; *Reich* 2010; *Meisel* 2011; *Berendes* et al. 2013). As regards the choice of an appropriate assessment instrument, *Ortega/Iberri-Shea* (2005, p. 40) noted that "(…) it is possible that certain tasks are inappropriate at early points of development (because they require knowledge of the L2 that is not in place yet) or for the initial waves of data collection (because the participants are too young in the beginning of the study to be able to understand the task construction)". These considerations further beg the questions whether one and the same instrument can be used to measure and compare language competencies in mono- and multilinguals, and which linguistic subdomains should be involved.

3 Language competencies at the beginning of the school career

For the present research purpose, an understanding of the nature of language competencies as well as a grounded knowledge on the trajectories of language development over time are needed, with a particular focus on the initial phase of schooling and differences in the patterns of language development in mono- and multilinguals. For a given language, acquisition stages themselves but not the rate at which children progress through them are assumed to be very similar across children (cf. *Mitchell/Myles* 2004). Whereas the different stages of language development are well-documented for toddlers and younger children (cf. *Reich* 2010), there is a shortage of studies that focus on the identification of relevant milestones for older speakers (cf. *Berman* 2004; *Tolchinsky* 2004; *Nippold* 2007). At any rate, though, there is evidence to suggest that the entry into and the initial phase of schooling is associated with both quantitative and qualitative changes in language competencies both in mono- and multilingual children (cf. *Berendes* et al. 2013). Arguments that have been brought forward in this vein relate to changes in children's language input towards a more school-related and conceptually written register (cf. *Gogolin/Lange* 2011) and a focus on alphabetization and literacy-based processes (cf. *Grießhaber* 2012). In particular, changes are expected in more complex language forms, such as sophisticated textual cohesion and syntactical forms (cf. *Nippold* 2007; *Held* 2009).

As regards domain-specific differences in the language competencies in mono- and multilinguals in the beginning of their educational careers, several studies have pointed to particularly strong discrepancies in language domains that are increasingly relevant in educational contexts where written language plays an important role, such as syntax, vocabulary and verbal morphology (cf. *Nippold* 2007; *Chlosta/Schäfer* 2008; *Ahrenholz* 2010; *Grießhaber* 2012). Complex syntactical forms as a means of establishing textual coherence as well as differentiated and technical vocabulary are typical language domains that are school-related and often acquired at a later phase of language development, especially in multilingual children (cf. *Berman* 2004; *Diehl* et al. 2000). Following the observation of a comparatively strong increase in the areas of verbal morphology and collocation forms in sentences in multilingual children in this phase, there is also evidence for a convergence in these domain-specific competencies sometimes as early as second grade depending on the precise language area under consideration (cf. *Grießhaber* 2003; *Dehn* 2006; *Herwartz-Emden* et al. 2008).

4 Research hypotheses

Based on the consideration that the initial phase of schooling may imply both a quantitative and qualitative change in language competencies (cf. *Berendes* et al. 2013), we expect to observe an increase in the overall test score from the first measurement point, when the participants were 6 and 7 years of age, to the second wave of data collection, which was conducted 9 to 15 months later. Further, we expect different developmental patterns in different language domains.

Based on the assumption that language development occurs in similar stages in mono- and multilingual children, but that the latter may reach certain milestones later than their monolingual peers, we expect to observe significantly lower test scores for multilingual children in the first wave but a convergence in the scores achieved by mono- and multilingual children over time. Based on empirical findings on language development in multilinguals (e.g., *Grießhaber* 2003; *Dehn* 2006; *Herwartz-Emden* et al. 2008), we expect this convergence to largely result from a comparatively strong increase in the school-related language domains syntax, vocabulary and verbal morphology in multilingual children.

5 Methodological approach

5.1 Study context and sample

The present study uses data from a pilot study for a proposed "LiMA-Language Development Panel Study[1], which investigates the language development of children and young adolescents in two data collection waves (2011 and 2012). The study involves three age cohorts (6/7, 11 and 15-year olds) that comprise monolingual Germans and persons with a Russian, Turkish and Vietnamese language background. A random sample based on register data was selected according to two criteria: (a) year of birth and (b) at least one parent with first or second nationality Russian, Turkish or Vietnamese and/or place of birth in one of the former states of the former Soviet Union, Turkey or Vietnam. To increase the sample size, this initial sample was complemented by snowballing methods (about 40 percent in the cohort of 6/7-year-olds). Data was collected in households both from the target subjects and their parents.

The present study uses data from the German language assessment task HAVAS 5, which was conducted in the cohort of 6/7-year olds (N=81). The youngest participant was almost 6 years of age and the oldest participant almost 7.5 years old at the first point of data collection in 2011 (mean age 6.6 years, SD=0.38). A second measurement was conducted 9 to 15 months later (mean time between data collection points 0.96 years (SD=0.12). The sample includes 52 girls (47%) and 58 boys (53%). While only 61 percent of children attended primary school in the first wave of data collection, this applies to all children in the second measurement point.[2]

5.2 The instrument "HAVAS 5"

Children's language proficiency in German was measured using the standardized instrument *HAVAS 5 – Hamburger Verfahren zur Analyse des Sprachstands Fünfjähriger* (cf. *Reich/*

Roth 2004). The tool elicits oral narrations based on a series of six pictures and captures productive oral language competencies in different domains. The instrument is particularly suited for the assessment of language competencies in multilingual children (cf. *Reich/Roth* 2004) as it captures several phenomena that are typical for second language acquisition (e.g., code-switching and phases of German verbal morphology). The linguistic analysis is based on the method of profile analysis (cf. *Clahsen* 1986), which aims at gaining detailed insight into different aspects of individual language development. The German version of HAVAS comprises the following language subdomains (cf. *Reich/Roth* 2004):

A. *Task accomplishment*: Overall narrative ability, assessed by the extent to which children refer to all actors and actions in the story. According to *Reich/Roth* (2004), this domain can be understood as a general cognitive measure for task engagement.
B. *Communication situation*: Pragmatic and interactional knowledge, including communicative features such as initiative, continuity and clarity of pronunciation, aspects of dealing with language difficulties as well as the use of the family language (i.e., code switching).
C. *Vocabulary*: Every single full verb used is marked once (types), including modal verbs. Auxiliary verbs are not counted.
D. *Verbal forms and position in the sentence*: Five stages of acquisition of German verbal morphology and positioning of verbs in different clause types (based on research on second language acquisition, cf. *Clahsen* 1991), ranging from simple verbs to complex verbal positioning in subordinate clauses. The category encompasses aspects of verbal morphology, such as the conjugation of different tenses, as well as syntactical aspects that derive from the collocation of verbs within complex sentences. The highest phase is marked.
E. *Sentence connectors*: Indicator for syntactical knowledge, reflecting the use of different types of clause connecting elements in subordinate clauses (e.g., consecutive, relative, final, conditional). This category also captures five phases of language acquisition (cf. *Clahsen* 1991).

The data was recorded with the software Audacity and transcribed using GAT conventions (cf. *Selting* et al. 1998); subsequent coding procedures were conducted by trained raters. Only bilingual interviewers were used in consideration that co-ethnic interviewers have been shown to be more successful both in winning migrants for survey participation as well as in administering testing procedures to them (cf. *Baykara-Krumme* 2010). All interviewers were trained intensively before each data collection wave and had several opportunities to analyze authentic interview examples before going into the field.[3]

6 Results

The first part of the analysis describes the mean changes in the test scores achieved by children in the first and second measurement point for the full sample and for the groups of mono- and multilinguals separately. Taking into consideration possible differences in the developmental patterns across children in general, and across mono- and multilinguals specifically, the second part provides information on individual patterns of language change. To take into account the potential existence of interviewer effects that may affect

the assessment of children's language development, the third part provides first results of qualitative data analyses.

6.1 Mean changes in mono- and multilingual children

Table 1 contains a summary of the mean test scores achieved in the first and second wave for the full sample, regarding both the overall test score and the scores the participants achieved in the different subdomains of the language test. Each subscore (subdomains A-E) represents the number of points achieved relative to the highest score that can theoretically be achieved and is reported in percent. The overall test score was calculated as the mean of the relative subscores (α=. 0.8 in wave one, α=. 0.6 in wave two).

Table 1. Mean changes in language test scores, full sample

Variable	Mean score wave 1	Mean score wave 2	Difference (in perc. points)
Overall score	60,02	65,33	5,31***
	(1.536)	(1.253)	(1.544)
Subscores			
A: Task accomplishment	52,81	63,85	11,04***
	(2.027)	(1.444)	(1.937)
B: Communication situation	80,79	80,79	0,00
	(1.575)	(1.513)	(1.798)
C: Vocabulary	38,61	38,55	-0,06
	(1.439)	(1.596)	(1.935)
D: Verbal forms and position	79,75	82,72	2,96
	(2.497)	(1.748)	(2.679)
E: Sentence connectors	48,15	60,74	12,59***
	(2.694)	(3.092)	(3.608)

Note: N=162/n=81; paired t-test, two-sided; *** p < 0.001; ** p < 0.01; *p < 0.05; std. errors in parentheses.

In line with the hypothesis that the initial phase of schooling will positively affect language development (cf. *Berendes* et al. 2013), the data points to a significant increase in the overall test score of about five percentage points from the first to the second wave. The separate consideration of the scores achieved in each subdomain further suggests that this increase is primarily attributable to a comparatively large improvement in the dimension task accomplishment (A) and in the use of sentence connectors (E).

Table 2 provides information on the overall test scores achieved by mono- and multilinguals in the first and second wave of data collection. The third column displays the mean differences in the test scores achieved by each group in the first and the second wave. The third row for each category displays the differences in the test scores achieved by mono- and multilingual children in each measurement point. As expected, the data points to significantly lower test scores in the group of multilingual children in the first wave. This finding not only applies to the overall test score but to every single subdomain with the exception of the dimension communication situation (B). Also, the data supports the notion of a comparatively strong improvement in multilingual children in several subdomains and hence of a convergence in the language competencies across the groups over

time. For multilingual children, a significant increase cannot only be observed for the overall test score but also in the dimensions task accomplishment (A), verbal forms and position in the sentence (D) and sentence connectors (E). For their monolingual peers, significant differences can neither be observed in the overall test score nor in any of the language subdomains with the exception of the dimension task accomplishment (A).

Table 2. Mean changes in language test scores, mono- and multilingual children

Variable	Mean score wave 1	Mean score wave 2	Difference w2 - w1 (in perc. points)
Overall score			
Monolinguals	68,70	70,30	1,60
	(1.452)	(1.853)	(2.508)
Multilinguals	56,58	63,36	6,78***
	(1.889)	(1.52)	(1.892)
	12,12***	6,94*	
	(3.144)	(2.685)	
A: Task accomplishment			
Monolinguals	61,71	71,01	9,30**
	(2.837)	(2.387)	(3.53)
Multilinguals	49,28	61,02	11,73***
	(2.459)	(1.648)	(2.327)
	12,43**	10,00**	
	(3.544)	(3.02)	
B: Communication situation			
Monolinguals	83,42	88,32	4,89
	(2.102)	(1.856)	(2.078)
Multilinguals	79,74	77,80	-1,94
	(2.03)	(1.846)	(2.332)
	3,68	10,51**	
	(3.49)	(3.162)	
C: Vocabulary			
Monolinguals	47,07	44,35	-2,72
	(1.705)	(3.77)	(4.479)
Multilinguals	35,26	36,25	0,99
	(1.708)	(1.579)	(2.049)
	11,81***	8,10*	
	(2.464)	(3.444)	
D: Verbal forms and position			
Monolinguals	87,83	80,00	-7,83
	(3.717)	(3.327)	(4.988)
Multilinguals	76,55	83,79	7,24**
	(3.078)	(2.055)	(3.02)
	11,27*	-3,79	
	(5.426)	(3.878)	
E: Sentence connectors			
Monolinguals	63,48	67,83	4,35
	(4.473)	(4.827)	(6.282)
Multilinguals	42,07	57,93	15,86***
	(2.98)	(3.832)	(4.338)
	21,41***	9,90	
	(5.508)	(5.7)	

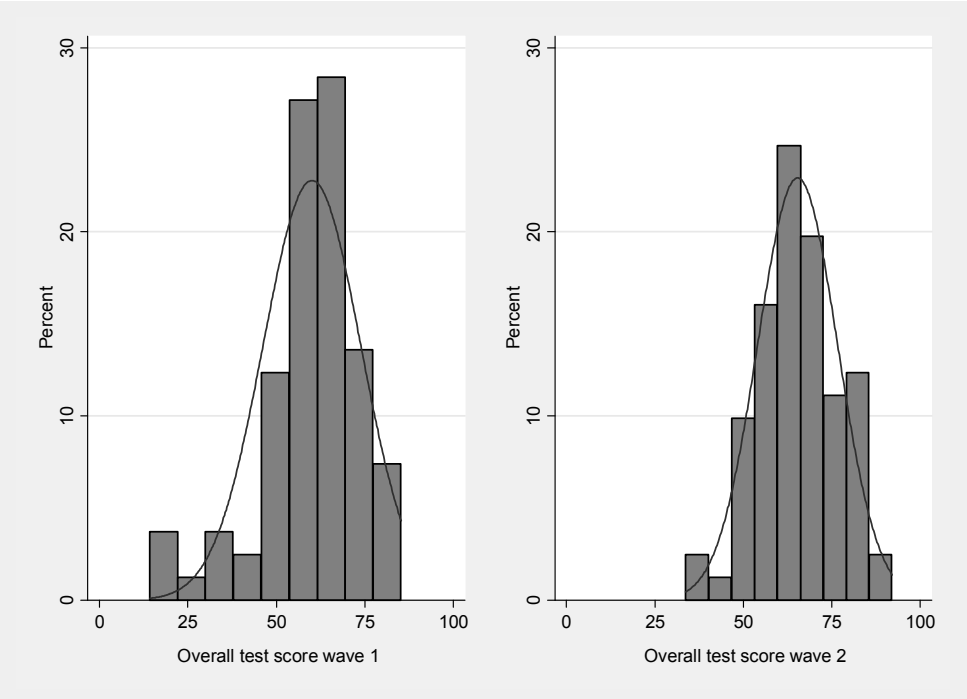
Note: N=162/n=81; two-sided t-test; *** $p < 0.001$; ** $p < 0.01$; *$p < 0.05$; std. errors in parentheses.

Further, the data points to different patterns of convergence across the two groups in the different language domains. Significant differences remain in the categories task accomplishment (A) and vocabulary (C), and an increase in the achievement gap can be observed for the scores achieved in the dimension communication situation (B). Conversely, the data points to a particularly strong convergence in the dimensions verbal forms and position (D) and sentence connectors (E), where no significant differences can be observed between mono- and multilinguals in the second measurement point. In fact, as regards the dimension verbal forms and position (D), a reversed pattern can be observed over time (although, the difference in wave two is not significant at any common level).

6.2 Individual patterns of change

Figure 1 presents two histograms that provide information on the overall test score the participants achieved in the first and second wave of data collection[4] (full sample, with a normal density added to the graph). As expected, the increase in the overall test score observed above appears to be largely attributable to improvements in children who scored comparatively low in the first wave. Counter to the hypothesis of an improvement in children's language competencies over time, however, the data also points to a decrease in the overall test score in the case of several children. The latter observation appears to particularly apply to children who scored comparatively high in the first measurement point.

Figure 1. Individual changes, full sample



To obtain further insight into this matter, figure 2 displays a histogram that represents the differences in the overall test scores across the two waves (i.e., the score achieved in the second wave minus the score achieved in the first wave, so that positive (negative) values indicate an increase (a decrease) over time). While positive changes can be observed for the majority of children (49 children, 60.5%), the data points to a substantial share whose test scores in the second wave fall short of those achieved in the first wave (31 children, 38.3%). The figure also clearly shows that changes in the overall test score appear to be random in some cases only, and points to considerable changes in the case of other children. Further, these patterns can be seen to be of different nature in the group of mono- and multilingual children (figure 3). In line with the observation of a comparatively high mean increase in the overall test score in multilingual children, the data points to an increase in the test score from the first to the second wave for the great majority of children from this group (67.2 percent). A reverse pattern can be observed in the group of monolinguals, where the majority (56.5 percent) scored lower in the second measurement point.

Figure 2. Differences within individuals, full sample

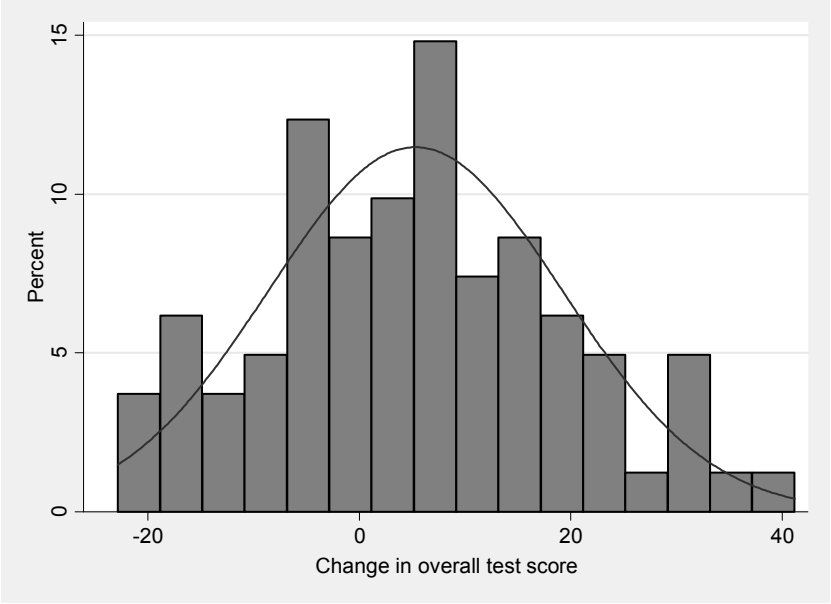
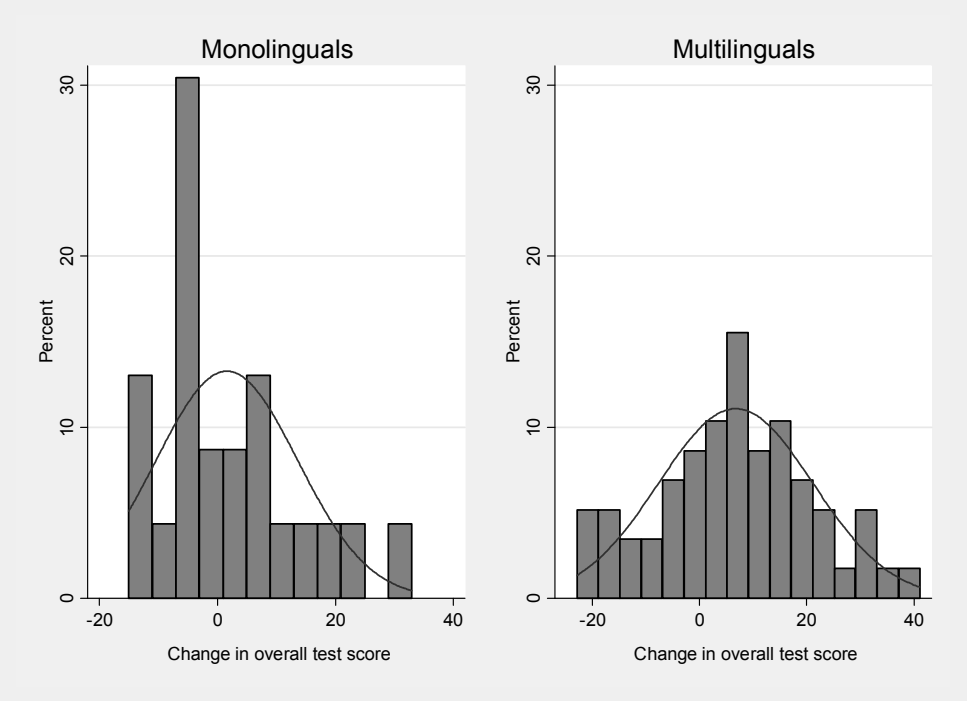


Figure 3. Differences within individuals, mono- and multilingual children



The observed stagnation in children's language development in certain subdomains may have several explanations. These include methodological concerns, such as the small sample size and the not normally distributed data (which make it difficult to identify statistical significance)[5], the occurrence of ceiling effects that may conceal language development (100 percent were scored in the first wave in the categories communication situation (B), verbal forms and positions (D) and sentence connectors (E), test effects that may result from the repeated use of the same instrument (such as boredom and lower levels of motivation in the second measurement point) and interviewer effects (that result from aspects such as the use of different question techniques, cf. *Ortega/Iberri-Shea* 2005; *Bachman* 2007). Unfortunately, the present data provides limited opportunities to (directly) investigate these considerations in more detail. To obtain further insight into the reasons for which lower test scores were measured for a considerable share of the sample in the second data collection wave, and to assess the extent to which the comparatively large increase in the test scores in multilinguals can indeed be explained by theoretical considerations, the following section provides first results of a qualitative analysis that identifies difficulties which specifically relate to variations in level of interviewer proficiency across the two measurement points and across mono- and multilingual children.

6.3 Qualitative excursion

To obtain first insight into possible explanations for the patterns of language change observed above, several interviews were reanalyzed to investigate the interviewers' proficiency, specifically with regard to the dimensions script adherence and communication style, while collecting the HAVAS data[6]. The sample for the qualitative analysis includes the 20 interviews where the largest decrease in the overall test score (8 monolingual and 12 multilingual children) and the 20 interviews where the strongest improvements were observed from the first to the second wave of data collection (4 monolingual and 16 multilingual children). The results point to comparatively strong variations in the interviewers' proficiency both (1) across the first and second wave and (2) across mono- and multilingual children.

As concerns the extent to which the interviewers adhered to the interview script, particularly strong deviations were observed in terms of a reduced amount of impulse questions the interviewers were instructed to pose during the assessment task. More specifically, the data points to a smaller number of impulse questions that were posed in the second measurement point among children whose test scores decreased, but to a comparatively stable pattern across both waves in the case of children whose scores improved over time. A specific example is the prompt to narrate the story a second time after the first try. Among children whose scores decreased over time, this step was neglected entirely in the case of one interview only in the first wave but in as many as 9 interviews in the second wave. Conversely, the question was not posed in two interviews only in the second measurement point among children whose test scores increased over time. Further, the data indicates that this mistake occurred much more often in interviews with monolingual children.

As regards variations in the communication style adopted by the interviewers, deviations from the script in terms of missing feedback signals to encourage the participants were observed in one case only in the first wave but in as many as 10 interviews in the second wave among children whose test scores decreased over time. Conversely, the feedback was provided in the majority of 17 (wave one) and 18 (wave two) cases among children whose scores increased over time. Also, the data indicates that the comparatively strong decrease in the test scores in monolingual children is partly attributable to interruptions by the interviewer. No such instances occurred in the first measurement point in either group. In the second point of data collection, participants were repeatedly interrupted in the case of four interviews with monolingual children but in none of the interviews that were conducted with multilinguals.

7 Summary and Discussion

The present article aimed to investigate the oral language development in mono- and multilingual children at the beginning of their educational careers, and to obtain insight into the difficulties that are related to the assessment of productive oral language competencies in general and in a longitudinal perspective specifically. In line with the hypothesis that the initial phase of schooling implies a quantitative and qualitative change in language competencies (cf. *Berendes* et al. 2013), the analysis revealed a significant increase in the overall test score for the full sample. Also, the results support the proposition of different

patterns of language development in different language subdomains (cf. *Karmiloff-Smith* 1995). For the full sample, significant positive mean changes were observed in subdomains that are thought to represent general measures of cognitive and linguistic engagement with the task (task accomplishment) on the one hand, and that capture the use of complex syntactical forms (sentence connectors) on the other hand. Due to the advanced pragmatic and interactive knowledge observed for the majority of children in the first wave, the stagnation in children's ability to cope with the communication situation is not surprising. The stagnation in children's verbal vocabulary was not expected but is in line with *Nippold* (2007) and *Tolchinski* (2004), who argue that the phase of language development following school entry displays a comparatively strong growth in other domains, such as syntactical knowledge.

Also, several findings in the course of the analysis are supportive of the hypothesis of divergent patterns of language development in mono- and multilingual children. As expected, monolinguals were shown to significantly outperform their multilingual peers in the first wave of data collection both with regard to the overall test score and the majority of the language subdomains that were assessed using HAVAS 5. In line with the literature (e.g., *Grießhaber* 2003; *Dehn* 2006; *Herwartz-Emden* et al. 2008), the convergence in the scores across the two groups over time is primarily attributable to the comparatively strong improvement in the scores achieved by multilingual children. For monolingual children, a significant improvement was observed in the dimension task accomplishment only, which indicates a cognitive growth in the engagement with the task but not necessarily linguistic development. Conversely, significant and positive changes were observed in several further language subdomains in the group of multilinguals. As expected, a particularly strong improvement was observed in the use of sentence connectors and the category verbal forms and positions in the sentence.

Yet, several methodological aspects were discussed above that may affect the assessment of children's language competencies and hence, at least partly, account for the different developmental patterns observed for mono- and multilinguals. The occurrence of test effects such as boredom and lower levels of motivation in the second measurement point is likely not least due to the repeated use of the same instrument (cf. *Ortega/Iberri-Shea* 2005; *Bachmann* 2007) and due to the fact that the HAVAS instrument was initially developed to capture language competencies in multilingual children aged 5 to 7 years (cf. *Reich/Roth* 2004). Given the higher initial language competencies of monolinguals in the first measurement point, these effects may affect this group particularly strongly and partly account for their lesser improvement.

The data does not allow to investigate these considerations in more detail. Yet, in line with results from other empirical studies on oral language testing (e.g., *Reeman/Alas/Liiv* 2013), the qualitative analysis pointed to the occurrence of substantial interviewer effects that relate both to variations in communication style and the extent of script adherence. Apart from difficulties that relate to the small sample size, these effects may contribute to the stagnation, or even decrease, in the scores achieved in several language subdomains over time. Further, the qualitative analysis suggests that the comparatively strong increase in the test scores among multilinguals may not merely be attributable to their lower initial language competencies, but partly result from variations in the interviewer proficiency.

Even though the present study does not provide conclusive evidence on the patterns of language development in mono- and multilingual children, the results give rise to sev-

eral methodological and practical implications with regard to the measurement of oral language competencies in general and over time specifically.

First, the observation of differences in the patterns of language development both across linguistic subdomains and between mono- and multilinguals clearly points to the importance of using instruments that involve several linguistic domains (cf. *Jude/Klieme* 2007). While the HAVAS 5 instrument appears to be suited to identify language domains that follow different patterns of language change, the results suggest the further development of the instrument to include additional linguistic subdomains and/or the adaptation of existing categories to avoid the occurrence of ceiling effects to better discriminate higher achievers.

Second, the results point to the need for the development of assessment categories and measurement techniques that are less vulnerable to interviewer effects. For instance, the assessment of language competencies in the domains verbal forms and position (D) and sentence connectors (E) by means of five categories that are thought to reflect different developmental stages is particularly likely to be affected by aspects such as variations in the way questions are posed and in the feedback provided to encourage the participants.

At the practical level, the findings above clearly call for the provision of systematic feedback to interviewers both during each and between data collection wave(s). Also, the results strongly suggest that the regular qualitative analyses of random samples and consequent interviewer feedback based on the results should become a part of quality measures in quantitative studies.

The considerations that the assessment of oral language data is related to comparatively strong difficulties in capturing children's actual linguistic competence as compared to receptive testing, but that the latter provides data on a reduced set of the language areas identified by *Ehlich, Bredel* and *Reich's* theoretical model (2008), further emphasize the need for an early reflection on the advantages and constraints of a given instrument in light of its theoretical model. Given the wide variety of test effects that are potentially inherent to the measurement of (oral) language competencies, the results clearly point to the need for a larger number of measurement points to allow for the identification of trends in children's language development. Also, in view of comparatively high levels of heterogeneity of multilingual children with regard to dimensions such as their migration and educational biographies, their socio-demographic position and language typological aspects, a closer examination of groups of immigrants with similar language backgrounds is certainly of relevance for further longitudinal studies on patterns of language change.

Notes

1 LiMA is the acronym for Linguistic Diversity Management in Urban Areas and is an interdisciplinary research cluster of the University of Hamburg (2009-2013), which aims at investigating multilingualism from a resource-oriented perspective (www.lima.uni-hamburg.de).

2 Only those children were included who participated in the test in both measurement points. Russian background: *N*=29; Turkish background: *N*=17; Vietnamese background: *N*=12; monolingual German: *N*=23. No different patterns were observed for the distribution of the age, sex, grade attendance and the time between the two measurement points in the groups of mono- and multilingual children.

3 The training was based on the guidelines for interviewing with HAVAS 5 (cf. *Reich/Roth* 2004), which include numerous quality criteria for interviewer behaviour during oral interviewing (e.g., the instrument is to be applied in the form of a structured oral interview with a sequence of six prompt

questions, oral behaviour of interviewers such as the provision of encouraging but not correcting feedback, avoidance of interruptions).

4 The height of the bars is scaled in percent, so that each bar's height is equal to the share of children in the respective category.

5 A more detailed analysis that takes into account the clustered structure of the data was not conducted. As the tests conducted below may not appropriately identify statistical significance due to the comparatively small sample size, especially when treating mono- and multilingual children as separate groups, and the not normally distributed data, the present results may be understood as an indication for existing developmental patterns but not as definite evidence.

6 The categories for this analysis were derived from the literature on interviewer effects during oral testing and on the guidelines of the HAVAS instrument (cf. *Reich/Roth* 2004).

References

Ahrenholz, B. (2010): Bildungssprache im Sachunterricht der Grundschule. In: *Ahrenholz, B.* (Ed.): Fachunterricht und Deutsch als Zweitsprache. – Tübingen, pp. 15-36.

Bachmann, L. F. (2007): What is the Construct? The Dialectic of Abilities and Contexts in Defining Constructs in Language Assessments. In: *Fox, J./Wesche, M./Bayliss, D./Cheng, L./Turner, C. E./ Doe, C.* (Eds.): Language Testing Reconsidered. – Ottawa, pp. 41-71.

Baykara-Krumme, H. (2010): Interviewereffekte in Bevölkerungsumfragen. Ein Beitrag zur Erklärung des Teilnahme-und Antwortverhaltens von Migranten. Arbeitspapiere des Beziehungs- und Familienpanels, 19. Online: http://www.pairfam.de/uploads/tx_sibibtex/arbeitspapier_19p.pdf, last accessed: 30.05.2014.

Berendes, K./Weinert, S./Zimmermann, S./Artelt, C. (2013): Assessing Language Indicators Across the Lifespan within the German National Educational Panel Study (NEPS). Journal for Educational Research Online, 5, 2, pp. 15-49.

Berman, R. A. (Ed.) (2004): Language Development across Childhood and Adolescence. – Amsterdam.

Brown, A. (2003): Interviewer Variation and Construction of Speaking Proficiency. Language Learning, 20, 1, pp. 1-25.

Brown, A./Lumley, T. (1997): Interviewer Variability in Specific-purpose Language Performance Tests. In: *Kohonen, V./Huhta, A./Kurki-Suonio, L./Luoma, S.* (Eds.): Current Developments and Alternatives in Language Assessment: Proceedings of LTRC 96. – Jyväskylä, pp. 137-150.

Chlosta, C./Schäfer, A. (2008): Deutsch als Zweitsprache im Fachunterricht. In: *Ahrenholz, B./Oomen-Welke, I.* (Eds.): Deutsch als Zweitsprache. – Baltmannsweiler, pp. 280-297.

Clahsen, H. (1986): Die Profilanalyse. – Berlin.

Clahsen, H. (1991): Child Language and Developmental Dysphasia. – Amsterdam.

De Bot, K./Schrauf, R. W. (2009): Language Development over the Lifespan. – New York.

Dehn, M. (2006): Zeit für die Schrift: Lesen lernen und Schreiben können. – Berlin.

Diehl, E./Christen, H./Leuenberger, S./Pelvat, I./Studer, T. (Eds.) (2000): Grammatikunterricht: Alles für der Katz? Untersuchungen zum Zweitsprachenerwerb Deutsch. – Tübingen.

Eckhardt, A. G./Grgic, M./Leu, H. R. (2011): Vermessung der Kindheit im Rahmen von Sprachstandserhebungen? *Diskurs Kindheits- und Jugendforschung, 6,* pp. 263-280.

Ehlich, K. (2007): Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund. Bildungsforschung Band 11. – Bonn.

Ehlich, K./Bredel, U./Reich, H. H. (2008): Referenzrahmen zur altersspezifischen Sprachaneignung. Bildungsforschung Band 29, 1. – Bonn.

Gogolin, I./Lange, I. (2011): Bildungssprache und durchgängige Sprachbildung. In: *Fürstenau, S./Gomolka, M.* (Eds.): Migration und schulischer Wandel: Mehrsprachigkeit. – Wiesbaden, pp. 107-127.

Grießhaber, W. (2003): Einblicke in zweitsprachliche Schriftspracherwerbsprozesse. In: *Baumann, M./Ossner, J.* (Eds.): Diagnose und Schrift II: Schreibfähigkeiten. OBST 67/04, pp. 65-87.

Grießhaber, W. (2012): Zweitsprachkenntnisse und Literalität in sehr frühen Lernertexten. In: *Ahrenholz, B./Knapp, W.* (Eds.): Sprachstand erheben – Spracherwerb erforschen. – Stuttgart, pp. 155-171.

Held, U. (2009): Diagnostik schriftsprachlicher Kompetenzen in der Primarstufe. In: *Jeuk, S./Schmid-Barkow, I.* (Eds.): Differenzen diagnostizieren und Kompetenzen fördern im Deutschunterricht. – Stuttgart.

Herwartz-Emden, L./Braun, C./Heinze, A./Rudolph-Albert, F./Reiss, K. (2008): Geschlechtsspezifische Leistungsentwicklung von Kindern mit und ohne Migrationshintergrund im frühen Grundschulalter. Zeitschrift für Grundschulforschung 1, 2, pp. 13-28.

Jude, N./Klieme, E. (2007): Sprachliche Kompetenz aus Sicht der pädagogisch-psychologischen Diagnostik. In: *Klieme, E./Beck, B.* (Eds.): Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International). – Weinheim, pp. 9-22.

Karmiloff-Smith, A. (1995): Beyond Modularity: A Developmental Perspective on Cognitive Science. – Cambridge.

Malvern, D. D./Richards, B. J. (2002): Investigating Accommodation in Language Proficiency Interviews Using a New Measure of Lexical Diversity. Language Testing, 19, 1, pp. 85-104.

McNamara, T. F. (1996): Measuring Second Language Performance. – Harlow.

McNamara, T. F./Lumley, T. (1997): The Effect of Interlocutor and Assessment Mode Variables in Overseas Assessments of Speaking Skills in Occupational Settings. Language Testing, 14, 2, pp. 140-156.

Meisel, J. M. (2011): First and Second Language Acquisition: Parallels and Differences. – Cambridge.

Mitchell, R./Myles, F. (2004): Second Language Learning Theories. – New York.

Nippold, M. A. (2007): Later language development: School-age Children, Adolescents, and Young Adults (3rd Ed.). – Austin.

O'Loughlin, K. J. (1997): The Comparability of Direct and Semi-direct Speaking Tests: A Case Study. PhD Thesis, Department of Linguistics and Applied Linguistics, University of Melbourne.

Ortega, L./Iberri-Shea, G. (2005): Longitudinal Research in Second Language Acquisition: Recent Trends and Future Directions. Annual Review of Applied Linguistics, 25, pp. 26-45.

Reemann, E./Alas, E./Liiv, S. (2013): Interviewer Behaviour During Oral Proficiency Interviews: A gender perspective. Estonian Papers in Applied Linguistics, 9, pp. 209-227.

Reich, H. H. (2005): Auch die „Verfahren zur Sprachstandsanalyse bei Kindern und Jugendlichen mit Migrationshintergrund“ haben ihre Geschichte. In: *Gogolin, I./Neumann, U./Roth, H.-J.* (Eds.): Sprachdiagnostik bei Kindern und Jugendlichen mit Migrationshintergrund. – Münster, pp. 87-95.

Reich, H. H. (2010): Frühe Mehrsprachigkeit aus linguistischer Perspektive. – München.

Reich, H. H./Roth, H.-J. (2004): Hamburger Verfahren zur Sprachstandsanalyse Fünfjähriger: Auswertungsbogen und Auswertungshinweise. – Hamburg.

Ross, S. (1996): Formulae and Inter-interviewer Variation in Oral Proficiency Interview Discourse. Language Testing, 9, 2, pp. 3-16.

Ross, S./Berwick, R. (1992): The Discourse of Accomodation in Oral Proficiency Interviews. Studies in Second Language Acquisition, 14, 2, pp. 159-176.

Selting, M./Auer, P./Barden, B./Bergmann, J./Couper-Kuhlen, E./Günthner, S./Meier, C./Quasthoff, U./Schlobinski, P./Uhmann, S. (1998): Gesprächsanalytisches Transkriptionssystem (GAT). Linguistische Berichte, 173, pp. 91-122.

Shohamy, E. (1983): The Stability of Oral Proficiency Assessment on the Oral Interview Testing Procedures. Language Learning, 33, 4, pp. 527-540.

Tolchinsky, L. (2004): The Nature and Scope of Later Language Development. In: *Berman, R.* (Ed.): Language Development across Childhood and Adolescence. – Amsterdam, pp. 233-248.

Weinert, S./Ebert, S. (2013): Spracherwerb im Vorschulalter: Soziale Disparitäten und Einflussvariablen auf den Grammatikerwerb. Zeitschrift für Erziehungswissenschaft, 16, 2, pp. 303-332.